Coefficient of Determination

Let's use the ovarian cancer dataset from the "curatedOvarianData" package in R Bio-conductor. A detailed description is available here. First we must upload the data in R:

```
source("https://bioconductor.org/biocLite.R")
biocLite("curatedOvarianData")
library(curatedOvarianData)
source(system.file("extdata", "patientselection.config", package = "curatedOvarianData"))
set.seed(123)
```

Now, predict the expression of the gene in the first column using the expressions in columns 2, 3,. . . , 6. The data matrix can be obtained with the command `mat.gene = exprs(TCGA_eset)`.

```
data("TCGA_eset")
varLabels(TCGA_eset)
```

```
##  [1] "alt_sample_name"
##  [2] "unique_patient_ID"
##  [3] "sample_type"
##  [4] "histological_type"
##  [5] "primarysite"
##  [6] "arrayedsite"
##  [7] "summarygrade"
##  [8] "summarystage"
##  [9] "tumorstage"
## [10] "substage"
## [11] "grade"
## [12] "age_at_initial_pathologic_diagnosis"
## [13] "pltx"
## [14] "tax"
## [15] "neo"
## [16] "days_to_tumor_recurrence"
## [17] "recurrence_status"
## [18] "days_to_death"
## [19] "vital_status"
## [20] "os_binary"
## [21] "relapse_binary"
## [22] "site_of_tumor_first_recurrence"
## [23] "primary_therapy_outcome_success"
## [24] "debulking"
## [25] "percent_normal_cells"
## [26] "percent_stromal_cells"
## [27] "percent_tumor_cells"
## [28] "batch"
## [29] "flag"
## [30] "flag_notes"
## [31] "uncurated_author_metadata"
```

```
mat.gene = exprs(TCGA_eset)

y = mat.gene[1,]
x = t(mat.gene[2:6,])

dd = cbind(y, x)
dd = as.data.frame(na.omit(dd))
head(dd)
```

```
##                      y       A2M    A4GNT     AAAS     AACS    AADAC
## TCGA.20.0987 2.923522 10.353008 3.321405 4.608010 7.279213 4.605331
## TCGA.23.1031 3.052169 11.635772 3.666463 5.142133 7.048869 5.775611
## TCGA.24.0979 2.846371  7.954542 3.258038 5.025422 7.750161 3.846412
## TCGA.23.1117 3.002209  9.971500 3.596212 5.139928 6.206031 4.468379
## TCGA.23.1021 3.062993  8.971334 3.388706 5.256831 7.835422 4.415817
## TCGA.04.1337 2.974734  9.042876 3.269979 4.667723 6.763047 4.159804
```

```
tail(dd)
```

```
##                      y       A2M    A4GNT     AAAS     AACS    AADAC
## TCGA.24.1852 2.804569  7.952748 3.266449 5.032060 7.181784 3.780469
## TCGA.29.1692 2.993727  9.068691 3.542584 4.877798 7.580572 8.451190
## TCGA.13.1817 3.136917 10.198890 3.336109 4.709619 5.921329 4.546632
## TCGA.61.1916 2.965996  9.699037 3.405650 5.145519 8.360284 6.119431
## TCGA.29.1704 3.157896  8.336289 3.323166 4.957783 5.700931 4.066970
## TCGA.13.1819 2.934607  8.661028 3.307428 5.404119 6.553031 4.178920
```

Now fit a linear regression model and report the R^2 value.